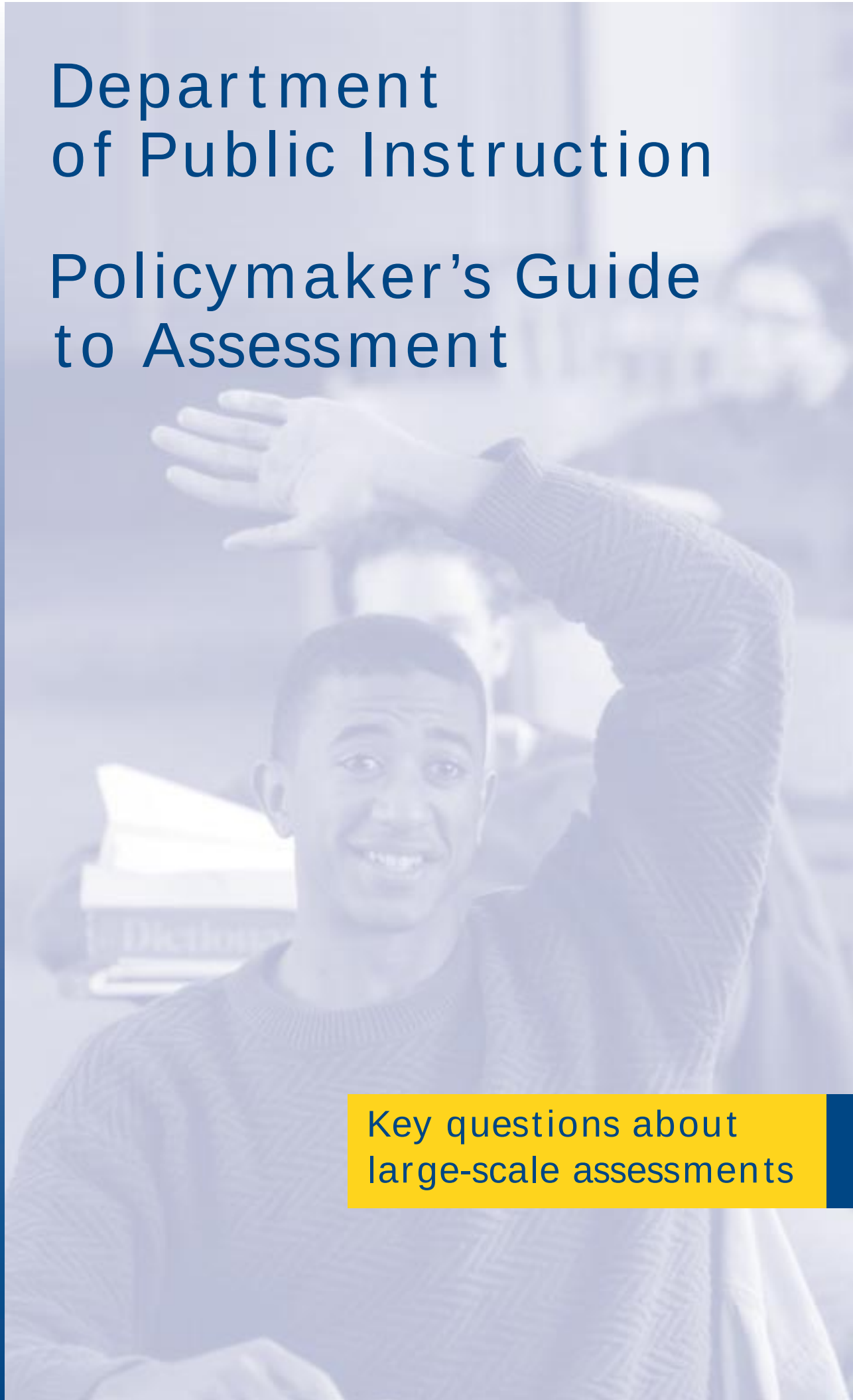
Department of Public Instruction

Policymaker's Guide to Assessment



Key questions about large-scale assessments

State of Wisconsin Department of Public Instruction

John T. Benson
State Superintendent

125 S. Webster St.
P.O. Box 7841
Madison, WI
53707-7841

1-800-441-4563
1-608-266-3390

www.dpi.state.wi.us.

Key questions about large-scale assessments

This Guide is designed to assist you in answering key questions about good test use and practice. The information is crucial to understanding and mandating tests. We hope the Guide is useful to you as you make decisions about Wisconsin's children and schools.

Question #1: What are tests used for in education?

Tests add value to the educational process, but are not a cure-all for the ills of the educational system. Tests are samples of behavior. When used properly, they allow us to make informed, insightful decisions about the status of our children and schools. When used improperly, tests may hinder the educational process.

Among the many purposes for testing in education, the most common are:

To diagnose students' academic situations

- Diagnostic tests are designed to identify strengths and weaknesses.
- The Wisconsin Reading Comprehension Test (given to all third graders in the state) is a diagnostic test. Its purpose is to diagnose reading problems, after which parents and teachers collaborate to address the problems.
- Diagnostic exams are very specific. They are used to make specific decisions about children and their abilities in very specific content areas.

To monitor students' academic achievement

- Achievement tests focus on a range of skills rather than on mastering a precise amount of content (like diagnostic tests).
- The Wisconsin Knowledge and Concepts Examinations (WKCE) at Grades 4, 8, and 10 are achievement tests.
- The purpose of achievement tests is to tell us about student knowledge and give us a degree of insight into how well the curriculum prepared the student. If students in a given school score particularly well on some aspect of an achievement test, there is a good chance that their curriculum and the manner of teaching the curriculum succeeded in preparing them for the test.

To certify whether students have met academic standards or not

- Certification tests focus on skills students are expected to know and demonstrate.
- The proposed Wisconsin High School Graduation Test (HSGT) is an example of a test used for certification based on the pre-determined Wisconsin Model Academic Standards.
- The primary purpose of a test used for certification is to distinguish between students who have met a pre-determined standard and those who have not. In the case of the HSGT, the purpose will be to distinguish between students who meet the assessed 12th grade academic standards and those who do not.

Question #2:

What are the benefits and limitations of large-scale tests?

In order to understand possible consequences when we choose or require large-scale tests, it is vital to examine the benefits and limitations of such testing.

Benefits of test use

Tests focus instruction on important content. Test results can help focus educational efforts on knowledge and skills that might otherwise go unlearned. This is true only for groups of students rather than for individuals. Cases where group results show that students are not mastering a particular area of content are signals that teachers need to make changes.

Valid, reliable, and well-understood tests provide teachers and students with an incentive for good teaching and strong learning.

Tests give incentives to teachers and students. Valid, reliable, and well-understood tests provide teachers and students with an incentive for good teaching and strong learning. Their mutual efforts are rewarded by high test scores and public recognition of good work.

Tests provide feedback on progress that can be translated into better education. Testing that is connected with the content taught provides essential feedback to teachers and students. Teachers learn more about the content and skills their students are mastering, and are better able to adjust their teaching strategies to meet student needs. Students receive feedback on their progress toward learning information and developing skills that have been defined as important.

Tests provide large-scale comparability. If large numbers of students take the same exam, Wisconsin policymakers, parents, teachers and the general public can compare test performance across schools, districts, and regions of the state. This comparability allows decision-makers to identify exemplary schools and schools that may require improvement.

Tests potentially provide evidence that assures the public and employers of a student's level of competency. This important benefit is dependent on the quality, validity, and reliability of the conclusions made possible by the test scores. (These terms are addressed in question #3.) However, many subjective, non test-related factors can influence how employers and the public perceive the value of a test in readying students to perform as expected. Assuring employers and the public that students have been readied to perform as expected—at a standard level of competence—is the goal of the High School Graduation Test. To achieve this, the conclusions drawn from the test must not only be valid and reliable, they must be perceived as valid and reliable.

(Question #2 Continued)

Assuring employers and the public that students have been readied to perform as expected is the goal of the High School Graduation Test.

LIMITATIONS OF TEST USE

Not all important school outcomes can be tested. Within the framework of the Model Academic Standards, several performance standards do not lend themselves well to large-scale, standardized testing.

For example, a fourth grade performance standard is "Orally communicate information, opinions, and ideas effectively to different audiences for a variety of purposes." This skill is important to master, but obviously can't be readily included on a written test. Classroom-based tests such as performance portfolios are better suited to determine whether students have mastered standards difficult to measure through large-scale assessment.

Student academic achievement would be quite limited if standards for student accountability were only those which could be measured by large-scale assessment.

Focusing solely on tests tends to narrow the curriculum. Educators may be motivated to narrow their curriculum to cover only those aspects of the Model Academic Standards which can be assessed through large-scale, standardized tests. Rewards and sanctions based solely on the results of such tests may encourage such narrowing. Media pressure will also reinforce narrowing. Media reports normally convey only the results of large-scale testing, not the complete "picture" of evaluative activities that focus on academic achievement within a school or district.

Negative emotional judgments of the test can undercut the test's acceptance as a valid standard of competence. Experience in other states demonstrates that sometimes the validity of the test as a predictor of competence can be undercut by a combination of three factors: anguish over taking the test, the test's level of difficulty, and the fallout from unexpected test scores.

Question #3:

How do you determine the quality of a test?

In developing educational tests, quality is judged on the basis of the validity and reliability of the conclusions drawn from test scores.

Please note: tests are not valid or invalid. It is the conclusions we draw from test scores that are valid or invalid.

VALIDITY

Validity is the most important characteristic of a quality test. A valid 10th grade science test makes it possible to draw adequate conclusions about a student's ability in 10th grade science. If it is genuinely valid, we can conclude that students who score well understand the content tested, and students who score poorly have not mastered 10th grade science.

Standardized test creation is an extremely collaborative, social process involving children, parents, educators, and the public at large.

Validity is demonstrated by the scope and quality of the content tested. The conclusions we draw from test scores are dependent upon the scope of the content tested. If the content is limited, it is important not to draw powerful conclusions from the test scores. If our 10th grade science test is missing key content taught in 10th grade, our conclusions are limited.

A test could be valid for one predictive purpose but less valid for a different purpose. The results from testing 10th grade science knowledge could be valid in predicting a student's ability to use the science knowledge. The results would not be valid for predicting future success in science, or the student's ability to teach science.

RELIABILITY

Tests must also be reliable. In the context of testing, "reliability" has much the same meaning as it does in the scientific method. The results must be repeatable and consistent. The reliability of a particular test is its ability to repeatedly produce consistent scores.

Reliability is particularly important for scores near a cut-off point. Known as "cut scores," these cut off points separate two levels of results. For example, a cut score separates the "proficient" level from the "basic" level. To identify student performance in these two categories and minimize miscategorization, a test must be trustworthy and consistent.

Including more test items is one way to increase the reliability of scores. It gives students more opportunity to reliably demonstrate their knowledge. The fewer test items, the less reliable the test.

Question #4:

How are standardized tests created?

It is a common misperception that standardized tests are created in locked rooms where test experts—distant from kids, teachers and teaching—churn out exams. Nothing can be further from the truth. Standardized test creation is an extremely collaborative, social process involving children, parents, educators, and the public at large.

Standardized tests are tests that have the same "standardized conditions" during administration and the same "standardized items" on the test.

The process of constructing a standardized test is time-consuming, complex, and involves many steps. The following is a partial list of the steps. Each of these steps involves extensive collaboration.

(1) Identify the purpose for which the test scores will be used.

(2) Define the content to be measured by the test.

(3) Choose the format of the test questions (multiple choice, short answer, essay, etc.), and decide what proportion of the test should be in each format.

(4) Write an initial group of test questions.

(5) Review and revise test questions.

(6) Set up try-out sessions to determine whether students understand the test questions, whether the questions are clear, etc.

(7) Field-test the draft test on a group of students who are similar to the ones who will ultimately be tested.

(8) Statistically analyze the results of the field test to determine which test questions are doing the best job of distinguishing between various levels of proficiency.

(9) Conduct reliability and validity studies of the test.

(10) Develop directions for administration and interpretation of the test scores.

This process is essential in developing quality, highly valued exams. Failure to go through this time-consuming, exhaustive process threatens the validity of the test. As a result, statewide-standardized measures typically take years in development.

A standardized test does not necessarily have just multiple-choice questions. Recently, tests using multiple-choice questions exclusively have been criticized as not testing "high-order" learning skills. While this is not necessarily true, multiple-choice questions can limit the types of responses children can give. Many new, professionally developed tests incorporate both multiple-choice, short answer, and essay questions. An essay exam can be standardized as long as it has the same "standardized conditions" during administration and the same "standardized items" on the test.

State of Wisconsin
Department of
Public Instruction

John T. Benson
State Superintendent

125 S. Webster St.
P.O. Box 7841
Madison, WI
53707-7841

1-800-441-4563
1-608-266-3390

www.dpi.state.wi.us.